

1

Corrigé du T.D. 3

Ex1

1) Séries Marginales

Séries Doubles

$X \backslash Y$	1-3	3-5	5-9	Série Marginale de X
1	4	8	<u>16</u>	28
2	6	12	24	42
3	<u>3</u>	6	12	21
4	2	4	8	14
Série Marginale de Y	15	30	<u>60</u>	<u>$n=105$</u>

2) Distribution Conditionnelle de Y lorsque X=4

Elle définit par le tableau suivant $f_{Y|X} = \frac{f_{ij}}{f_{i.}}$

Valeur de Y	1-3	3-5	5-9
fréquence	0,143	0,286	0,571

$$= \frac{m_{ij}}{m_{i.}}$$

3) 16 : est le nombre de salariés dont le salaire mensuel est dans la tranche [5-9[et possédant 1 seul enfant.

3 : est le nbre de salariés possédant 3 enfants et ayant un salaire dans la tranche [1-3[

4) Vérifier que les 2 variables X et Y sont indépendantes

1^{ère} manière : $\rho_{xy} = \frac{1}{n} (\sum x_i y_i) - \bar{x} \cdot \bar{y} = 0$

2^{ème} manière : $r_{xy} = \frac{\rho_{xy}}{s_x \cdot s_y} = 0$

2

Dans ce cas les deux droites de régression
sont orthogonales

5) Variances marginales

$$\bar{x} = 2,2$$

$$s_x^2 = 0,96$$

$$\bar{y} = 5,43$$

$$s_y^2 = 3,67$$

Exercice 9

Valeur de x	10	20	35	50	70	90	110	130
Valeur de y	5	3,75	2,75	2,5	1,75	1,5	0,8	0,5

1) Les 2 caractères X et Y sont-ils corrélés?

On calcule la covariance de la pêne double s_{xy}

par :
$$s_{xy} = \left(\frac{1}{n} \sum_{i=1}^8 x_i y_i \right) - \bar{x} \cdot \bar{y}$$

$$= 90,21875 - (64,375 \times 2,25625)$$

$$= -55,027 \neq 0$$

Donc X et Y sont corrélés

ou bien on calcule le coefficient de
corrélation linéaire r_{xy} par :

$$r_{xy} = \frac{s_{xy}}{s_x \cdot s_y} = -0,9485$$

- 2) Donner la droite de Meyer en prenant comme
G₁ point moyen de 4 premiers points et
G₂ " " " 4 derniers points

$$G_1 \begin{pmatrix} x_1 \\ y_1 \end{pmatrix} \quad G_2 \begin{pmatrix} x_2 \\ y_2 \end{pmatrix}$$

3

$$x_1 = \frac{10 + 20 + 35 + 50}{4} = 28,75$$

$$y_1 = \frac{5 + 3,75 + 2,75 + 2,5}{4} = 3,4375$$

$$x_2 = \frac{70 + 90 + 110 + 130}{4} = 100$$

$$y_2 = \frac{1,75 + 1,25 + 0,8495}{4} = 1,075$$

Drate de Meyer et celle passant par les deux points G_1 et G_2

On obtient a et b par

$$a = -0,03315 \quad b = 4,39$$

$$\boxed{y = -0,03315x + 4,39} \quad \begin{matrix} \text{pour } x = 140 \\ \hat{y} = -0,251 \\ \text{modèle incompatible} \end{matrix}$$

3) Calculer le coefficient de détermination r^2

$$r = -0,9485 \quad r^2 = 0,8996$$

r^2 varie entre 0 et 1 ; il permet de juger la qualité d'une régression linéaire simple ; il mesure l'adéquation entre le modèle linéaire explicatif et les données observées.

• $r^2 \approx 0$ le modèle explicatif linéaire n'est pas adapté au modèle des points observés
 $r^2 = 1$ est le cas où le modèle est adapté au modèle des données observées

4) Modèle linéaire de régression de y par x

$$y - \bar{y} = \frac{s_{xy}}{s_x^2} (x - \bar{x})$$

$$y = \hat{a}x + \hat{b}$$

$$\hat{a} = \frac{s_{xy}}{s_x^2} = -0,03367$$

$$\hat{b} = \bar{y} - \frac{s_{xy}}{s_x^2} \cdot \bar{x} = 0,08874$$

$$y = \frac{s_{xy}}{s_x^2} \cdot x + \bar{y} - \frac{s_{xy}}{s_x^2} \bar{x}$$

$$= -0,03367x + 0,08874$$

pour $x = 140$ $y = -4,62506$

Modèle incompatible!

5) $y = a \ln(x) + b$

Calcul de a et b

1°/ on pose $u = \ln(x)$ $y = au + b$

$$\hat{a} = \frac{s_{\ln(x)y}}{s_{\ln(x)}^2}$$

$$\hat{b} = \bar{y} - \hat{a} \cdot \bar{x}$$

On trouve : $\hat{a} = -1,7221$

$$\hat{b} = 8,9472$$

pour $x = 140$ $y = 0,4369$

Modèle compatible qui ajuste au mieux les données.